

Penny Statistics¹

U.S. pennies minted after 1982 have a Zn core with a Cu overlayer. Prior to 1982, pennies were made of brass, with a uniform composition (95 wt% Cu / 5 wt% Zn). In 1982, both the heavier brass coins and the lighter zinc coins were made. In this experiment, you will use penny mass data collected in previous labs to answer the following questions: (1) Do pennies from different years have the same mass? (2) Do the masses follow a Gaussian distribution?

Supplied Data

Each student will be supplied an Excel file that contains mass data for pennies minted in the years between 1975 and 2000, including both brass and zinc pennies from 1982, sorted by year and by ascending mass, and an unsorted file of the combined mass data between 1983 and 2000.

If “Data Analysis” is not an option in the Excel Data ribbon, click the File menu, and select “Options” and “Add-Ins”. Select “Analysis ToolPak”, click GO, check the box next to “Analysis ToolPak” and then OK to load the ToolPak.

Confidence Intervals and t Test

Compute the average and standard deviation for each of the columns, writing the results for each column in Rows 40 and 41. Check your results by comparing your values for year 2000 to the “true” values of 2.494 and 0.008125, respectively.

Copy the data for years 1975 to 1989 and paste them into the sheet corresponding to an unlabeled data tab. Label the tab with the range of years. Select the two years in which the zinc coins (years ≥ 1982) have the highest and lowest average masses (look beyond the third decimal place if necessary). Calculate the value of F for the two data sets, and determine if they are drawn from populations with the same population standard deviation (σ), thus making the standard deviations for the two samples statistically equivalent. Next, open a new data tab in Excel and follow the example on pages 79-80 in your textbook with the appropriate t -test to determine if the means of the two data sets are statistically equivalent at the 95% confidence level.

Repeat the calculations for the two years in which the brass coins have the highest and lowest average masses.

Distribution of Masses

Unsorted mass data for years 1983 to 2000 can be found in the data tab labeled with those years. Copy the data and paste into a different column, then sort the data from low mass to high mass. Find the mean (\bar{x}), median, and standard deviation (s) for this data set.

Next, divide the data into 0.01-g intervals (e.g., 2.420 to 2.430 g) between 2.420 and 2.610 (a total of 20 intervals; you may find the “Count” function useful here) and prepare an appropriately labeled bar graph, like that shown in Figure 1. Use the center value of your range (e.g. 2.425) for your x-value. For random (Gaussian) data, only 3 out of 1000 measurements should lie outside of $\bar{x} \pm 3s$. Indicate which bars (if any) lie beyond $\pm 3s$. In Figure 1 two bars at the right are outside of $\bar{x} \pm 3s$.

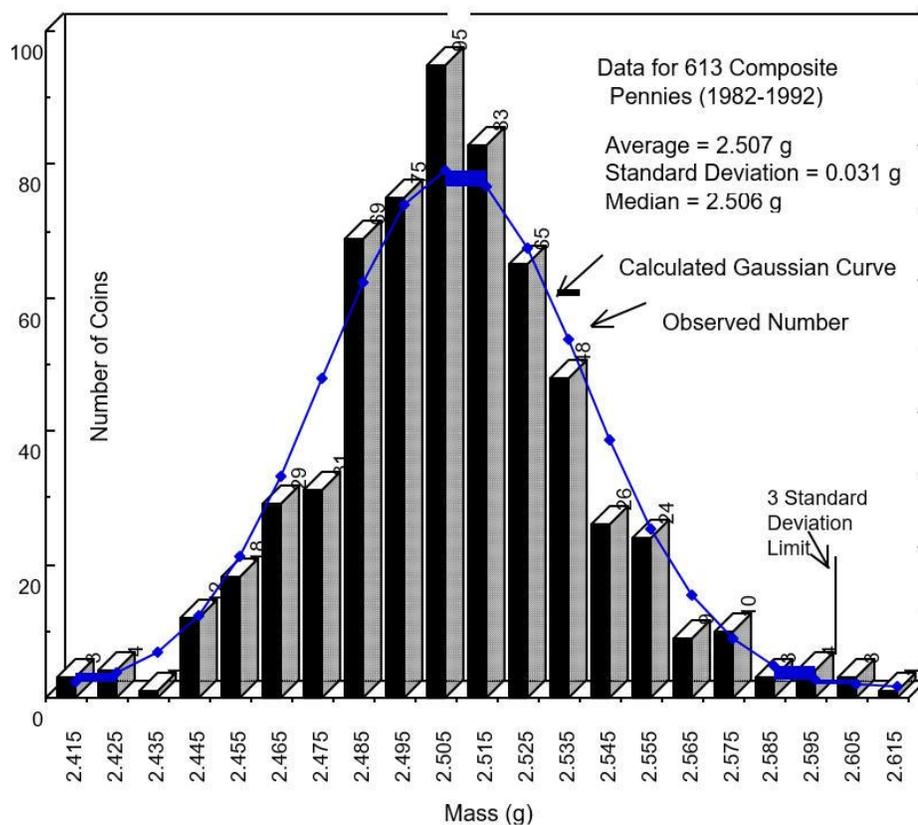


Figure 1. Distribution of penny masses from the years 1982 to 1992 measured in Dan's house by Jimmy Kuszniir and Doug Harris in December 1992.

The smooth Gaussian curve superimposed on the data in Figure 1 was calculated from the formula

$$y_{\text{calc}} = \frac{\text{number of coins}/100}{s \sqrt{2\pi}} e^{-(x - \bar{x})^2/2s^2}$$

Dividing the number of coins by 100 considers that the data points were grouped into ranges of 0.010 g each. Add a third column beside your range and frequency data and in the top cell insert this formula from with your specific values for n, s, and \bar{x} to allow the calculation of the *expected* number of coins in each mass range if a Gaussian distribution exists for the data (“lock” the locations of mean and std. dev. in the formula). Calculate by hand the first y_{calc} to make sure your equation is entered correctly. An example equation might look like:

$$=((613/100)/(\$G\$3*2.5066))*EXP(-((I2-\$G\$2)^2/(2*\$G\$3^2)))$$

Adding the scatterplot of this data to your existing plot will require the use of a “combo” plot combining a bar graph and a scatterplot.

Least-Squares Analysis: Do Pennies Have the Same Mass Each Year?

Copy the data for individual years 1983 to 1992 (first data tab) and paste them into a sheet corresponding to an unused data tab. Prepare a graph like Figure 2 in which the ordinate (y-axis) is the mass of pennies minted each year between 1983 and 1992, and the abscissa (x-axis) is the year. For simplicity, let 1983 be year 1, 1984 be year 2, and so on. You will have multiple entries for the abscissa corresponding to each ordinate (163 pennies total).

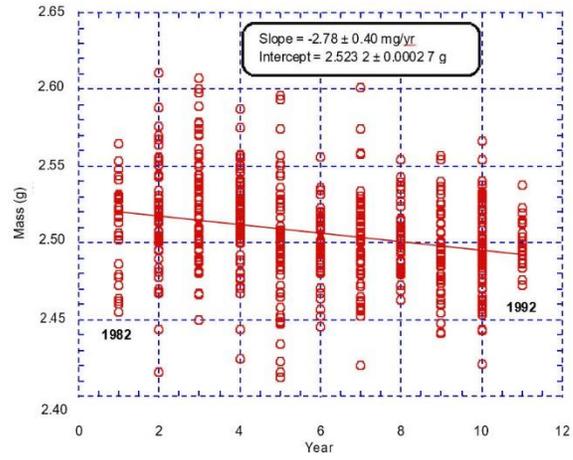


Figure 2. Penny mass versus year for 612 coins. Because the slope of the least-squares line is *significantly* less than 0, we conclude that the average mass of older pennies is greater than the average mass of newer pennies.

If the mass of a penny increases systematically from year to year, then the least-squares line through the data will have a positive slope. If the mass decreases, the slope will be negative. If the mass is constant, the slope will be 0. Even if the mass is really constant, your selection of coins is random and the slope is not exactly 0.

We want to know whether the slope is *significantly* different from zero. Use the LINEST function in Excel (see Section 4.9 in your textbook if you need a refresher) and calculate the slope (m) and intercept (b) of the best straight line through all points and find the uncertainties in slope (u_m) and intercept (u_b).

Use Student's t to find the 95% confidence interval for the slope:

$$\text{confidence interval for slope} = m \pm tu_m$$

where Student's t is for $n - 2$ degrees of freedom. For example, if you have $n = 300$ pennies, $n - 2 = 298$, and it would be reasonable to use the value of $t (= 1.960)$ at the bottom of the Table 4-4 in your textbook for $n = \infty$. If the least-squares slope is $m \pm u_m = -2.78 \pm 0.40$ mg/year, then the 95% confidence interval is $m \pm tu_m = -2.78 \pm (1.960)(0.40) = -2.78 \pm 0.78$ mg/year.

The 95% confidence interval is $-2.78 \pm 0.78 = -3.56$ to -2.00 mg/year. We are 95% confident that the true slope is in this range and is, therefore, *not* 0. We conclude that older zinc pennies are heavier than newer zinc pennies.

Does the Distribution of Masses Follow a Gaussian Distribution?

Now we carry out a χ^2 test (pronounced “ki squared”) to see if the observed distribution (the bars in Figure 1, 1983 – 2000) agrees with the Gaussian curve (scatterplot). The statistic χ^2 is given by

$$\chi^2 = \sum \frac{(y_{\text{obs}} - y_{\text{calc}})^2}{y_{\text{calc}}}$$

where y_{obs} is the height of a bar on the chart, y_{calc} is the ordinate of the Gaussian curve (Equation 2), and the sum extends over all bars in the graph. Calculations for the data in Figure 1 (not your data!) are shown in Table 1.

Table 1. Calculation of χ^2 for Figure 1

| Mass (g) | Observed number of coins | Calculated ordinate of Gaussian curve | | |
|----------|--------------------------|---------------------------------------|---|--|
| (x) | (y_{obs}) | (y_{calc}) | $y_{\text{obs}} - y_{\text{calc}}$ | $\frac{(y_{\text{obs}} - y_{\text{calc}})^2}{y_{\text{calc}}}$ |
| 2.415 | 3 | 1.060 | 1.940 | 3.550 |
| 2.425 | 4 | 2.566 | 1.434 | 0.801 |
| 2.435 | 1 | 5.611 | -4.611 | 3.789 |
| 2.445 | 12 | 11.083 | 0.917 | 0.076 |
| 2.455 | 18 | 19.776 | -1.776 | 0.159 |
| 2.465 | 29 | 31.875 | -2.875 | 0.259 |
| 2.475 | 31 | 46.409 | -15.409 | 5.116 |
| 2.485 | 69 | 61.039 | 7.961 | 1.038 |
| 2.495 | 75 | 72.519 | 2.481 | 0.085 |
| 2.505 | 95 | 77.829 | 17.171 | 3.788 |
| 2.515 | 83 | 75.453 | 7.547 | 0.755 |
| 2.525 | 65 | 66.077 | -1.077 | 0.176 |
| 2.535 | 48 | 52.272 | -4.272 | 0.349 |
| 2.545 | 26 | 37.354 | -11.354 | 3.451 |
| 2.555 | 24 | 24.112 | -0.112 | 0.001 |
| 2.565 | 9 | 14.060 | -5.060 | 1.821 |
| 2.575 | 10 | 7.406 | 2.594 | 0.909 |
| 2.585 | 3 | 3.524 | -0.524 | 0.078 |
| 2.595 | 4 | 1.515 | 2.485 | 4.076 |
| 2.605 | 3 | 0.588 | 2.412 | 9.894 |
| 2.615 | 1 | 0.206 | 0.794 | 3.060 |
| | | | χ^2 (all 21 points) = | 43.231 |
| | | | χ^2 (19 points — omitting bottom two points) = | 30.277 |

At the bottom of Table 1 we see that χ^2 for all 21 bars is 43.231. In Table 2 we find a critical value of 31.4 for 20 degrees of freedom (degrees of freedom = one less than number of categories). Because χ^2 from Equation 3 exceeds the critical value, we conclude that *the distribution is not Gaussian*.

It would be reasonable to omit the smallest bars at the edge of the graph from the calculation of χ^2 because these bars contain the fewest observations but make large contributions to χ^2 . Suppose that we reject bars lying >3 standard deviations from the mean. This removes the two bars at the right side of Figure 1 which give the last two entries in Table 1. Omitting these two points gives $\chi^2 = 30.277$, which is still greater than the critical value of

28.9 for 18 degrees of freedom in Table 2. Our conclusion is that at the 95% confidence level the observed distribution in Figure 1 is not quite Gaussian. It is possible that exceptionally light coins are nicked and exceptionally heavy coins are dirty or corroded. You would need to inspect these coins to verify this hypothesis.

Table 2. Critical values of χ^2 that will be exceeded in 5% of experiments*

| Degrees of freedom | Critical value | Degrees of freedom | Critical value | Degrees of freedom | Critical value |
|--------------------|----------------|--------------------|----------------|--------------------|----------------|
| 1 | 3.84 | 11 | 19.7 | 21 | 32.7 |
| 2 | 5.99 | 12 | 21.0 | 22 | 33.9 |
| 3 | 7.81 | 13 | 22.4 | 23 | 35.2 |
| 4 | 9.49 | 14 | 23.7 | 24 | 36.4 |
| 5 | 11.1 | 15 | 25.0 | 25 | 37.7 |
| 6 | 12.6 | 16 | 26.3 | 26 | 38.9 |
| 7 | 14.1 | 17 | 27.6 | 27 | 40.1 |
| 8 | 15.5 | 18 | 28.9 | 28 | 41.3 |
| 9 | 16.9 | 19 | 30.1 | 29 | 42.6 |
| 10 | 18.3 | 20 | 31.4 | 30 | 43.8 |

**Example:* The value of χ^2 from 15 observations is 17.2. This value is less than 23.7 listed for 14 (= 15 - 1) degrees of freedom. Because χ^2 does not exceed the critical value, the observed distribution is consistent with the theoretical distribution.

χ^2 for your data

Copy the three columns of data you used to construct Figure 1, and paste the values into a clean data tab. Construct a table like Table 1 using your data, and determine whether or not the distribution of your observed values between 1983 and 2000 is Gaussian.

Reporting Your Results

1. Complete the data sheet below using your calculations for this experiment, scan it or photograph it to produce either a pdf or jpeg file.
2. Annotate each tab in your Excel file to make clear where each section of data resides (Least Squares, for example). Make sure the data page includes all the calculations and graphs made during the lab period, and that each is clearly labeled.
3. Submit the pdf or jpeg from 1) with your Excel file via email to rskelly@esu.edu.
Please include your last name as part of the file name for each file.

Experiments to Accompany Quantitative Chemical Analysis 9th Edition, Daniel C. Harris, available online at <http://www.macmillanlearning.com/Catalog/studentresources/qca9e>, pp. 21-28; accessed September 21, 2018.

Data Table for Penny Statistics

t-test:

For the year ≥ 1982 (zinc coins) with highest average mass:

Year _____ Avg. Mass _____ Std. Dev. _____ n = _____

For the year ≥ 1982 (zinc coins) with lowest average mass:

Year _____ Avg. Mass _____ Std. Dev. _____ n = _____

Comparison of means with *t*-test:

$F_{\text{calculated}} =$ _____

$F_{\text{table}} =$ _____ Conclusion? _____

$t_{\text{calculated}} =$ _____

$t_{\text{critical}} =$ _____

Is the difference significant? _____

For the year ≤ 1982 (brass coins) with highest average mass:

Year _____ Avg. Mass _____ Std. Dev. _____ n = _____

For the year ≤ 1982 (brass coins) with lowest average mass:

Year _____ Avg. Mass _____ Std. Dev. _____ n = _____

Comparison of means with *t*-test:

$F_{\text{calculated}} =$ _____

$F_{\text{table}} =$ _____ Conclusion? _____

$t_{\text{calculated}} =$ _____

$t_{\text{critical}} =$ _____ Is the difference significant? _____

Least-squares Analysis:

Prepare a graph analogous to Figure 2 and find the least-squares slope and intercept and their standard deviations.

$m \pm u_m =$ _____

$t_{95\% \text{confidence}} =$ _____ $t_{99\% \text{confidence}} =$ _____

95% confidence interval: $m \pm tu_m =$ _____. Interval include zero? _____

99% confidence interval: $m \pm tu_m =$ _____. Interval include zero? _____

Is there a systematic increase or decrease of penny mass with year? _____

χ^2 test:

Construct a table analogous to Table 1 to compute χ^2 for the mass data from 1983 to 2000 plotted in Figure 1, without deleting any values.

Computed value of $\chi^2 =$ _____ Degrees of freedom = _____

Critical value of $\chi^2 =$ _____

Do the data follow a Gaussian distribution? _____

Now omit data on the graph that are greater than 3 standard deviations away from mean and calculate a new value of χ^2 with the reduced data set.

Mean value of all 406 data points _____ Std. dev. of all 406 data points _____

Value of mean + 3 s _____ Number of data points omitted _____

Computed value of $\chi^2 =$ _____ Degrees of freedom = _____

Critical value of $\chi^2 =$ _____

Does the reduced data set follow a Gaussian distribution? _____

Directions for Combo Plot

With the bar graph of “Number of Coins” vs “Mass” selected:

- 1) Click values “funnel” at right. Choose “Select Data”, then “Add” under “Legend Entry (Series)”.
- 2) Add the calculated y-values using “Series Values” to highlight the column. Hit “OK”, followed by “OK”.
- 3) Right click in the graph window and choose “Change Chart Type”. Choose “Combo”.
- 4) Select “Clustered Column” for Series 1, and “Line” for Series 2 (no secondary axis).
- 5) Hit “OK”.
- 6) Click on the line of the scatterplot, and while still on the line, right mouse click to select “Change Series Chart Type”.
- 7) Change Series 2 to “Scatter with Smooth Lines and Markers”, then “OK”.